User Guide

Updated February 4th, 2010

Jason A. Coombs

Benjamin H. Letcher

Keith H. Nislow

Program In Organismic and Evolutionary Biology
University of Massachusetts, Amherst, MA 01003
jcoombs@cns.umass.edu

S.O. Conte Anadromous Fish Research Center
US Geological Survey/Leetown Science Center
Turners Falls, MA 01376 USA

Northern Research Station
US Forest Service
University of Massachusetts, Amherst, MA 01003 USA

Table of Contents

Introduction. 3
Citation. 4
Installation. 4
Getting Started. 5
Batch File Format 10
Sibship Reconstruction Programs 15

- Colony v1.2 16
- Colony v2.0 16
- Kingroup 17
- Kinalyzer 18
- Parentage 19
- Pedigree 20
- PRT 20

Parentage Assignment Programs 21

- Cervus 21
- Colony v2.0 23
- Famoz 24
- Famsphere 25
- Gimlet 26
- Newpat 26
- Papa 27
- Parente 28
- Pasos 30
- Pedapp 30
- Probmax 31
- WhichParents 31

PedAgree 32

Sibship Reconstruction Output 33
Parentage Assignment Output 36
Sibship-Parentage Reconstructed-Reconstructed Comparison 40
Example Data 45
Disclaimer and Feedback. 45
References. 46

Introduction

PedAgree is a program for the rapid comparison of reconstructed relationships against a known pedigree for evaluative purposes, and for comparison of reconstructed relationships against one another to assess congruence of differing algorithms. Specifically it was designed as a companion program to the simulation software PEDAGOG (Coombs *et al.* 2010a) to evaluate existing sibship reconstruction and parentage assignment programs. Theoretically PedAgree could be used with any dataset for which there is a known pedigree. The advantage of using it with PEDAGOG is that PEDAGOG allows the user to vary numerous parameters such as mating strategy, mate choice, fecundity, cohort size, capture probability, survival, migration rates, age error, sex error, known parent error, genotype error, mutation, linkage, and genotype scoring probability to name a few. All of these parameters are then output to an apparent genotype file that represents a realistic dataset. The advantage lies in the fact that PEDAGOG also outputs a true genotype file with parental information for each individual. This allows the user to evaluate a sibship reconstruction or parentage assignment program by using an input file created from the apparent genotypes file (either specified and made during the PEDAGOG simulation, or made post-hoc using the input file creation program CREATE (Coombs *et al.* 2008)) and comparing it to the known pedigree file. Additionally, PEDAGOG is able to produce a batch file for rapid analysis using PedAgree.

For sibship reconstruction, comparisons are made between the number and composition of assigned and true full-sib families. Three scenarios exist for comparison of family number: there are fewer assigned families than true families, there are more assigned families than true families, or the numbers of assigned and true families are equal. The first scenario indicates that assignment grouped some true families together, while the second scenario indicates that assignment split some families apart. The composition of each true family is recovered by listing all the assigned families and their proportion of individuals found in the true family. A score is given showing the ratio of individuals assigned to the correct family divided by total individuals.

For parentage assignment, comparisons are made between the assigned and true parents. Errors can arise when any of the following occur: 1) an incorrect individual was assigned when the true parent was sampled, 2) no individual was assigned when the true parent was sampled, or 3) an incorrect individual was assigned when the true parent was not sampled. Errors are broken down into situations where neither true parent was sampled, one true parent was sampled while one was not, or both true parents were sampled. For cases where an incorrect parent was assigned when the true parent was sampled, the error is further broken down by classifying the relationship between the two as either full-sib, half-sib, or unrelated. Finally, accuracy is assessed for instances where both parents were assigned regardless of whether the true parents were sampled or not.

Citation

To cite PedAgree please use the following:

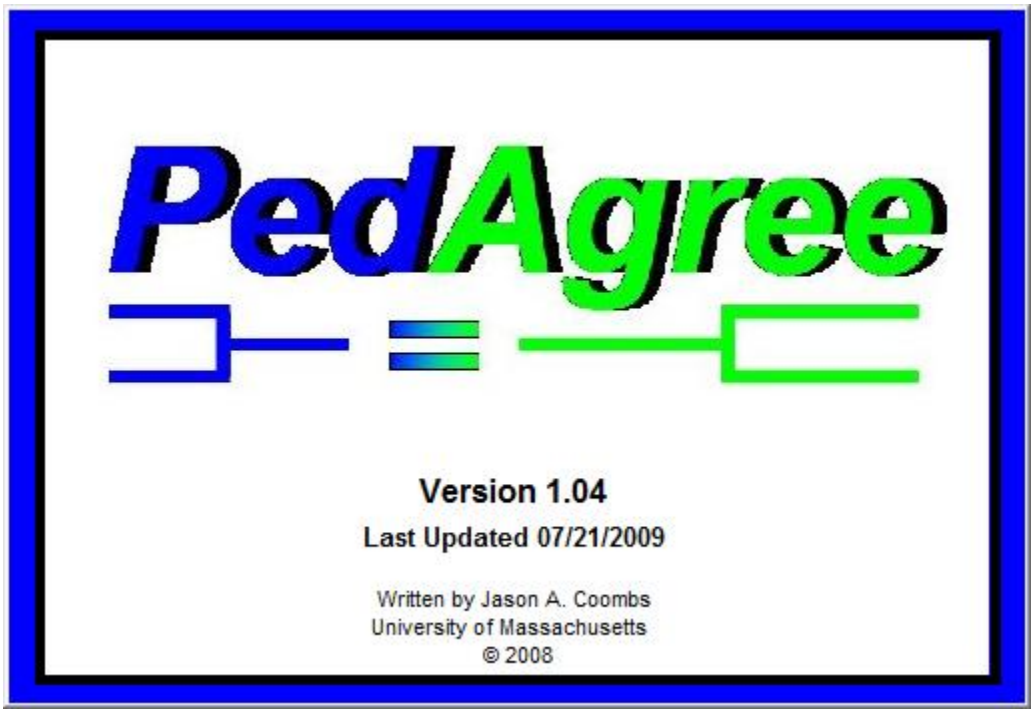
Coombs, J.A., Letcher, B.H., and Nislow, K.H. 2008. PedAgree: Software to quantify error and assess accuracy and congruence for genetically reconstructed pedigree relationships. Conservation Genetics Resources *In press*.

Installation

A zip file containing the program PedAgree, a user guide, and example data files is freely available for download from https://bcrc.bio.umass.edu/pedigreesoftware/.
The program is a stand-alone executable version that is fully operational once unzipped. Example files are included to provide data for experimentation with PedAgree in order for the user to gain familiarity.

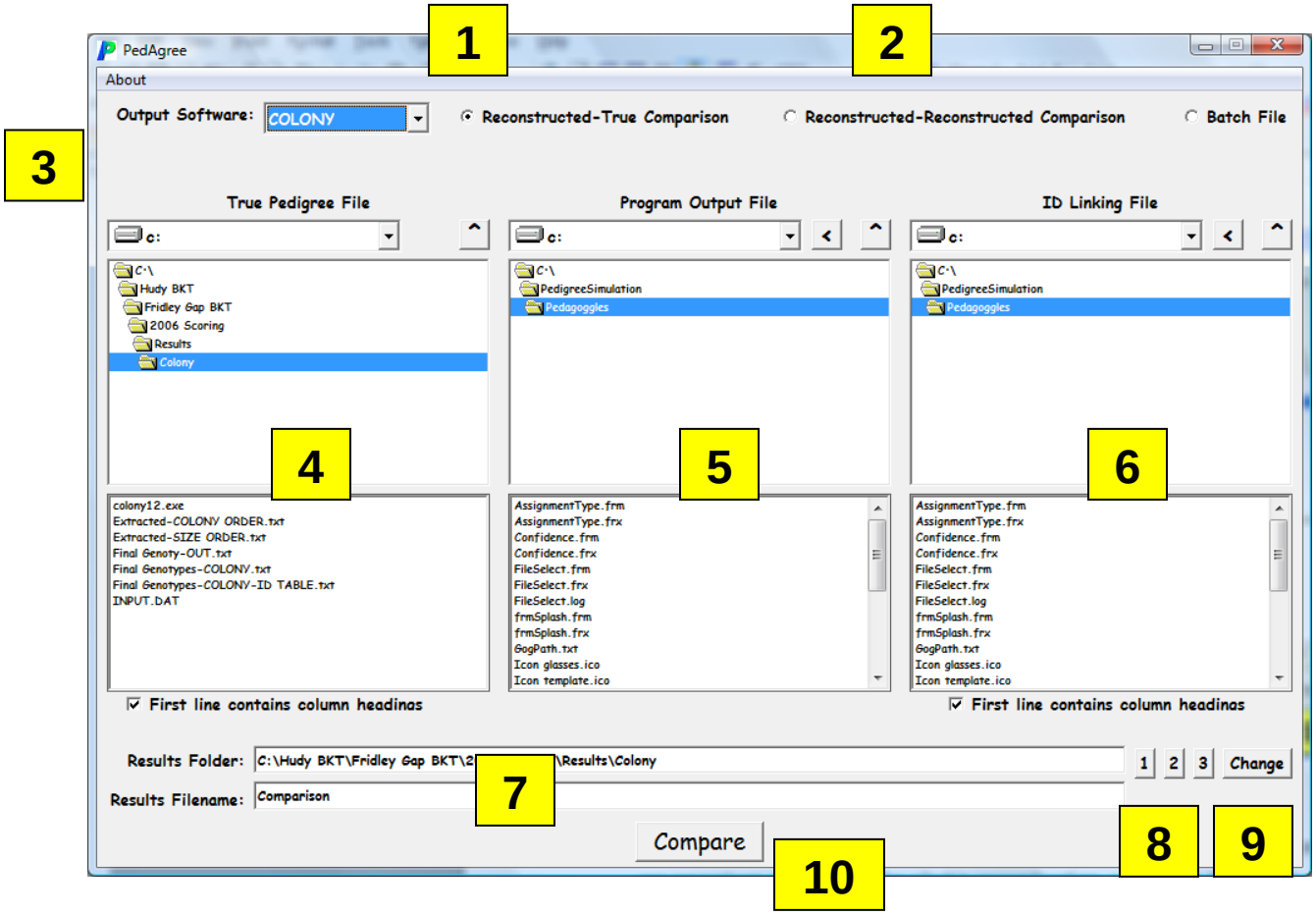
Getting Started

To start the program simply double click on the PedAgree icon or right click on the PedAgree icon and select Open. You will be momentarily presented with the following information screen:



This screen provides you with the program version number, and the date last updated for the executable you are running. You can return to this screen at any time by clicking on the ‘About’ menu option located on the main form.

Once the information screen departs you are presented with the following main screen:



1) The dropdown list allows you to choose which sibship reconstruction or parentage assignment program you wish to evaluate output for. Currently there are seventeen choices: five for sibship reconstruction, and twelve for parentage assignment. The Reconstructed-True Comparison and Batch File options require a true pedigree file, and an output file from the program selected. The Reconstructed-Reconstructed Comparison requires output files from the two programs involved. Additionally, certain programs require a linking file that relates generic labels with actual identifiers, or input files to account for which individuals were available for assignment.
2) The option buttons allow you to specify if you are going to compare a reconstructed output file against a true (known) pedigree file, compare two reconstructed output files, or run a batch file. Selection of either Reconstructed-True or Reconstructed-Reconstructed Comparisons means that the user must

specify all of the information mentioned in #1 above. Selection for a batch file means that the user must specify the file containing information for a series of output comparisons. Batch files can be created by the software PEDAGOG if specified. Making them by hand would most likely take longer then running each output individually.

3) Selecting the Reconstructed-Reconstructed Comparison option makes visible a tab option allowing for the selection of the two relationship reconstruction files to compare.
4) If the Reconstructed-True Comparison output option is selected then this region of the form specifies the location of the true or known pedigree file. At a minimum, the file itself must contain three columns; offspring, dam, and sire. They must be in that order and they must be separated by tabs. The file may contain other information in columns located to the right of these three columns, but this information will not be used by PedAgree. Simply select the drive and folder housing the file, then highlight the file in the file listbox. The button displaying the '^' character moves the current folder up one level. The checkbox underneath the file selection allows the user to specify whether or not the first line of the true pedigree file contains column headings.

   If the Reconstructed-Reconstructed Comparison output option is selected then this region of the form becomes disabled since there is no need for a true pedigree file.

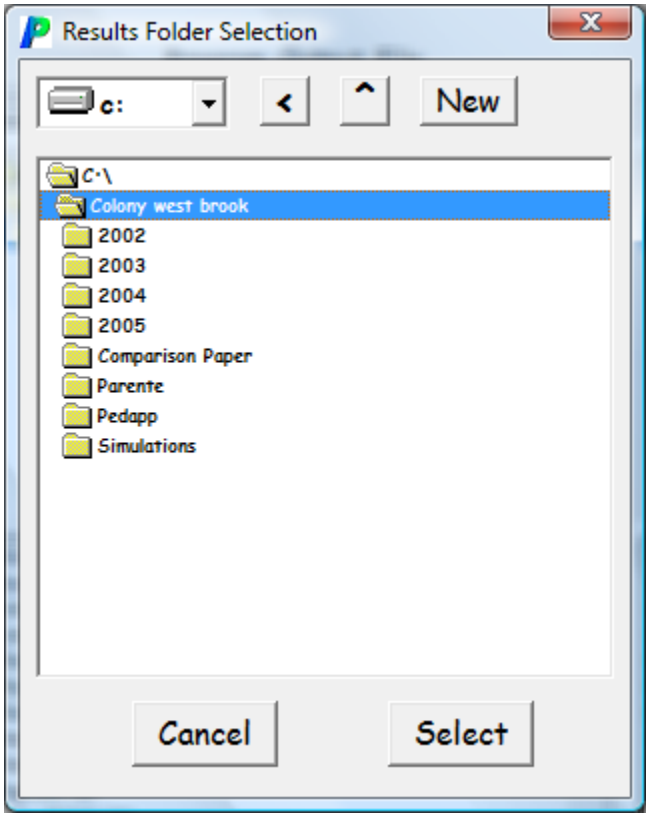
   If the Batch File option is selected then this region of the form specifies the location of the batch file. Simply select the drive and folder housing the file, then highlight the file itself in the file listbox. The button displaying the '^' character moves the current folder up one level.
5) This region of the form specifies the location of the output file produced by the program selected in step 1. Simply select the drive and folder housing the file, then highlight the file itself in the file listbox. The button displaying the '<' character changes the current folder to the folder selected in the true pedigree

directory. The button displaying the '^' character moves the current folder up one level.

6) This region of the form specifies the location of the Identifier linking file/candidate parent and offspring file associated with the program selected in step 1. Not all programs require this input. If the selected program does not require it, then this region will be disabled and a selection will be unable to be made.
   Several of the programs require the input of an ID linking file which links the generic identifier used by the program to the actual identifier used in the true pedigree file. This file must contain two columns of information separated by a tab. The two columns contain the identifier found in the true pedigree file and the generic identifier used by the sibship reconstruction or parentage assignment program.
   Four programs (CERVUS, FAMSPHERE, PASOS and PARENTE) require the selection of the input file or files used by the selected program in order to account for which individuals were available for assignment. Simply select the drive and folder housing the file(s), and then highlight the file(s) in the file listbox. To select multiple files press and hold the control key while clicking on the file name with the mouse. The button displaying the '<' character changes the current folder to the folder selected in the program output directory. The button displaying the '^' character moves the current folder up one level.
7) The results folder text box holds the pathway to the folder where you want the results to be output. The results filename textbox holds the name of what you wish to call the results file. You do not need to add an extension to the file, one will be added automatically. If the Reconstructed-True Comparison or Batch File option is selected then output files will have the name of the program producing the output and '–TRUE Details.txt' and '– Summary.txt' extensions. If the Reconstructed-Reconstructed Comparison option is selected then output files will have the name of the program producing the first output followed by the name of the program producing the second output followed by 'Details.txt' and 'Summary.txt'. If the comparison is between sibship and parentage output

sources a file with the first and second output program names followed by 'Pedigree.txt' is also created.

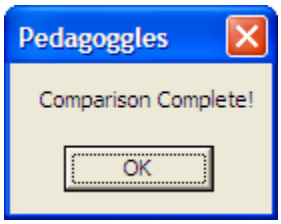
8) The command buttons labeled '1', '2', and '3' will set the results folder text box to the pathway for the selected folder of the true pedigree, program output, and ID linking/candidate parents and offspring respectively.
9) The command button labeled 'Change' allows the user to change the results folder pathway. Clicking this button opens up the folder selection form seen below:



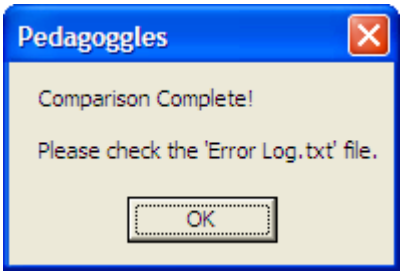
Simply select the drive and folder where you want the results file to be stored. The command button labeled '<' resets the pathway to what was selected when the form opened. The command button labeled '^' changes the pathway of the folder up one level. The command button labeled 'New' allows for the creation of a new folder. The new folder will be created in the folder currently selected. The command button labeled 'Cancel' closes the form without changing the pathway in the results folder text box. The command button labeled 'Select' closes the

form and changes the pathway in the results folder text box to that selected in the form.

10) The command button labeled ‘Compare’ runs the algorithm of the program that does the comparison of the specified files for all three options. When the program is finished it will present the following pop-up:



If you were running a batch file and there was an error encountered then you will receive the following popup:



This indicates that at least one of the comparisons specified in the batch file was unable to be run. Check the Error Log.txt file to determine which comparison generated the error. This file will be created in the same folder in which the batch file was located.

Batch File Format

Batch files are a way to run multiple sibship reconstruction and/or parentage assignment output comparisons using a single file. The format of the batch file input specifications can take one of three forms depending on which sibship reconstruction/parentage assignment program is being specified. There are explanations and examples of the three forms given below. Additionally, there is an example batch file titled ‘batch.txt’ located in the ‘Example Data’ folder. Lines inputting file location information may be specified in one of two ways:

1) The entire path of the file may be listed. An example line might be C:\Software\PedAgree\Demo\Parentage\output file.txt

2) If the folder housing the batch file is the root directory for the file to be specified then the line can begin with two dots followed by the remainder of the pathway. For example, if the batch file is located in the Demo file from the example in part 1 then an equivalent input line would be
..\ Parentage\output file.txt

Typically a user will not generate batch files by hand because it would take longer than running single output comparisons. Instead, batch files can be created while running simulations in the program PEDAGOG (Coombs *et al.* 2010a) and subsequently used by PedAgree.

Format 1

The first form consists of four lines, and is used for the following programs that do not require an ID linking or candidate parent and offspring file(s): COLONY_V2-SIB, COLONY_V2-PAR, KINGROUP, KINALYZER, PEDIGREE, GIMLET, PAPA, PROBMAX, and WHICHPARENTS.

Here is an example:

Papa
..\True Genotypes Trimmed.txt
..\Parentage Assignment\PAPA\allocation results.xls
..\Batch Results\Comparison

The first line specifies the program (Must match spelling on dropdown list).
The second line specifies the pathway to the true genotype file.
The third line specifies the pathway to the program output file.
The fourth line specifies the pathway and name of the file where results will be saved.

Format 2

The second form consists of five lines, and is used for the following programs that require an ID linking file: COLONY_V1, PRT, FAMOZ, and PEDAPP.

Here is an example:

Colony_V1
..\True Genotypes Trimmed.txt
..\Sibship Reconstruction\COLONY V1\output colony.txt
..\Sibship Reconstruction\COLONY V1\COLONY-ID TABLE.txt
..\Batch Results\Comparison

The first line specifies the program (Must match spelling on dropdown list).
The second line specifies the pathway to the true genotype file.
The third line specifies the pathway to the program output file.
The fourth line specifies the pathway to the ID linking file.
The fifth line specifies the pathway and name of the file where results will be saved.

Format 3
The last form consists of at least six lines, and is used for the following programs that require an ID linking file or a candidate parent and offspring file or files, and/or a probability or level of analysis: PARENTAGE, CERVUS, FAMSPHERE, NEWPAT, PASOS, and PARENTE.

NEWPAT does not require a linking or candidate parent and offspring file, but does require a value specifying whether the analysis is will be for maternity or paternity.

Here is an example for NEWPAT

Newpat
..\True Genotypes Trimmed.txt
..\Parentage Assignment\NEWPAT\RESULTS.TXT
0
..\Batch Results\Comparison

1

The first line specifies the program (Must match spelling on dropdown list).
The second line specifies the pathway to the true genotype file.
The third line specifies the pathway to the program output file.
The fourth line specifies the number of candidate parent and offspring files, which in the case of NEWPAT is always zero.
The fifth line specifies the pathway and name of the file where results will be saved.
The sixth line contains a value of either 0 or 1 indicating maternity and paternity analyses respectively.

See the NEWPAT program section for more details.

PARENTAGE and PARENTE both require a single file specifying an ID linking file and candidate parent and offspring file respectively, and an additional line specifying a probability level for analysis.

Here is an example for PARENTE:

Parente
..\True Genotypes Trimmed.txt
..\Parentage Assignment\PARENTE\pair.txt
1
..\Parentage Assignment\PARENTE\appa-PAR.txt
..\Batch Results\Comparison
0.25

The first line specifies the program.
The second line specifies the pathway to the true genotype file.
The third line specifies the pathway to the program output file.
The fourth line specifies the number of candidate parent and offspring files.

The fifth line specifies the pathway to the candidate parent and offspring file.
The sixth line specifies the pathway and name of the file where results will be saved.
The seventh line specifies the assigned parent probability for which to conduct the comparison analysis.

See the PARENTAGE or PARENTE program section for more details.

CERVUS, FAMSPHERE, and PASOS require the candidate offspring file and the candidate parents file (if sex is known then both the candidate mothers and candidate fathers files must be specified for CERVUS and PASOS). CERVUS requires an additional line containing an integer between 0 and 4 specifying the analysis level.

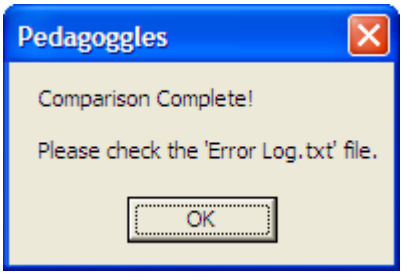
Here is an example for CERVUS:

Cervus
..\True Genotypes Trimmed.txt
..\Parentage Assignment\CERVUS\output cervus.csv
2
..\Parentage Assignment\CERVUS\CERVUS-OFFSPRING.txt
..\Parentage Assignment\CERVUS\CERVUS-PARENTS.txt
..\Batch Results\Comparison
3

The first line specifies the program (Must match spelling on dropdown list).
The second line specifies the pathway to the true genotype file.
The third line specifies the pathway to the program output file.
The fourth line specifies the number of candidate parent and offspring files.
The fifth line specifies the pathway to the first candidate parent and offspring file.
The sixth line specifies the pathway to the second candidate parent and offspring file.
The seventh line specifies the pathway and name of the file where results will be saved.

The eighth line specifies the assigned parent confidence level for which to conduct the comparison analysis. See the CERVUS program section for more details. FAMSPHERE and PASOS do not require the eighth line.

If there was an error encountered during the run, you will see the following popup upon completion:



The ‘Error Log.txt’ file will be located in the same folder as the selected batch file. Upon opening the file you will see a line for each error encountered. Below is an example of the contents found in an error log file.

2	CERVUS	Could not find true pedigree file
3	PARENTE	Could not find ID linking/parent-offspring input file

The first column refers to the ordered numerical location of the comparison generating the error. In the example both the second and third comparisons in the batch file generated errors.
The second column refers to the selected program of the comparison generating the error. In the example the programs of the error generating comparisons are CERVUS and PARENTE.
The third column explains what generated the error. In the example the CERVUS comparison could not locate the true pedigree file, while the PARENTE comparison could not locate the parent-offspring input file.

Sibship Reconstruction Programs

PedAgree currently makes comparisons for seven sibship reconstruction programs. All seven programs require the presence of a true pedigree file and an output file

generated by the selected program. Additionally some of the programs require an ID linking file to relate the generic identifiers used by the program to the actual identifiers used in the true pedigree file. For each of the seven programs a general overview of what the program does, additional requirements needed for PedAgree to make the comparison, the reference for the program, and a website from which the program can be downloaded are specified.

Colony v1.2

Overview

Colony assigns individuals sampled from a single generation of a population into full-sib families nested within half-sib families (colonies) using data on co-dominant genetic markers using a maximum likelihood method. It can be used in estimating full- and half-sib relationships, inferring mating systems (polygamous / monogamous) and reproductive skew in both diploid and haplo-diploid species, and reconstructing parental genotypes.

Additional Requirements

In order to make the comparison PedAgree requires an ID linking file.

Reference

(Wang 2004)

Download

http://www.zsl.org/science/research/software/colony,1154,AR.html

Colony v2.0

Overview

Colony is a computer program implementing a maximum likelihood method to assign/infer parentage and sibship among individuals using their multi-locus genotypes. It can be used in estimating full- and half-sib relationships, assigning

parentage, inferring mating systems (polygamous / monogamous) and reproductive skew in both diploid and haplo-diploid species.

In brief, the method (model) assumes a sample of individuals subdivided into 3 sub-samples: offspring (OFS), candidate males (CMS) and candidate females (CFS). Individuals in OFS are assigned (clustered) to K1 paternal and K2 maternal families (K1 and K2 unknown), individuals in CMS and CFS are assigned paternity and maternity to these K1 and K2 families. It is assumed that offspring are either full sibs (sharing both parents), half sibs (sharing only one of the two parents), or unrelated (sharing no parents), while candidates are unrelated and are either parents of or unrelated to the offspring. Markers are assumed to be in linkage equilibrium and Hardy-Weinberg equilibrium. Violation of these assumptions may lower the power of the analysis, but could be compensated by using more informative markers (Wang 2004). For example, the information about the sex and age of the sampled individuals might be unavailable. In such a case, each individual is allowed to appear in all 3 sub-samples and the sibship and parentage are still inferred satisfactorily in some cases.

Additional Requirements

For sibship analysis select the file with the extension '.BestFSFamily' extension

Reference

(Wang & Santure 2009)

Download

http://www.zsl.org/science/research/software/colony,1154,AR.html

Kingroup

Overview

KINGROUP's main objective is to test hypotheses of pedigree relationships between pairs of individuals using data from codominant, single-locus genetic markers.

Additional Requirements

After the analysis has run in KINGROUP, click anywhere on the results table and select all the output (control + A). Copy the output (control + C) and paste (control + V) it into a new notepad file and save it as your output file.

Reference

(Konovalov *et al.* 2004)

Download

http://code.google.com/p/kingroup/

Kinalyzer

Overview

Kinalyzer is a software suite developed to reconstruct sibling groups using genotypes from codominant markers such as microsatellites. Currently there are two algorithms available to reconstruct full-sibling groups for cases where parental genotypes are not available. Kinalyzer uses combinatorial optimization based on Mendelian inheritance rules to find the fewest number of sibling groups that contain all the individuals in the sample ('2-allele set cover'). Also available is a 'consensus' method that reconstructs sibgroups using subsets of loci and finds the consensus of these different solutions.

Additional Requirements

None

Reference

(Ashley et al. 2009)

Download

http://kinalyzer.cs.uic.edu/

Parentage

Overview

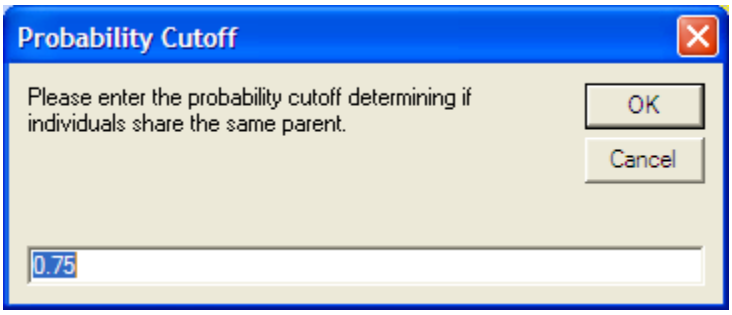
PARENTAGE was developed to draw inferences for the problem of shared maternity with an unknown number of fathers and no additional information except for allele frequencies in the breeding population.

Additional Requirements

Select the output file that has the '.paternity' extension.

In order to make the comparison PedAgree requires an ID linking file.

When running a PARENTAGE comparison in PedAgree, you will be prompted by the below form for a probability at which to determine if individuals share a same parent.



This value must be greater than or equal to 0 and less then or equal to 1. If specifying a PARENTAGE comparison in a batch file, the probability value is entered on the seventh line.

Reference

(Emery et al. 2001)

Download

http://www.staff.ncl.ac.uk/i.j.wilson/

Pedigree

Overview

The primary objective of PEDIGREE is to reconstruct the full pedigree in a group of individuals based on their genotype data with the complete absence of parental information. The program allows the user to potentially reconstruct the set of single generation relationships among individuals (i.e. which individuals are most likely full-sibs, half-sibs and unrelated). In addition, it allows generating the genotype of the unknown parents.

Additional Requirements

Once the run has completed go to the 'View Details' link under the 'Commands' column heading. For whichever run you wish to analyze, click on the 'Groups' link under the 'View' column heading. Click anywhere on the page and select all the output (control + A). Copy the output (control + C) and paste (control + V) it into a new notepad file and save it as your output file.

Reference

(Smith *et al.* 2001,Butler *et al.* 2004)

Download

http://herbinger.biology.dal.ca:5080/Pedigree/

PRT

Overview

Pedigree Reconstruction Tools (PRT) is specifically designed for a single generation of organisms, without parents present. The input is a table of DNA

markers, and the output is a partition of the organisms into putative sibling groups.

Additional Requirements

Before running the analysis in PRT, select and delete all data in the output window. Run the analysis in PRT. Once PRT has finished running click on the 'File' menu and select 'Save output'. Select this file for the program output file in PedAgree.

In order to make the comparison PedAgree requires an ID linking file.

Reference

(Almudevar & Field 1999)

Download

http://www.urmc.rochester.edu/biostat/people/faculty/almudevar.cfm

Parentage Assignment Programs

PedAgree currently makes comparisons for thirteen parentage assignment programs. All thirteen programs require the presence of a true pedigree file and an output file generated by the selected program. Additionally some of the programs require an ID linking file to relate the generic identifiers used by the program to the actual identifiers used in the true pedigree file, or a candidate parent and offspring file or files to account for the availability of potential parents. For each of the thirteen programs a general overview of what the program does, additional requirements needed for PedAgree to make the comparison, the reference for the program, and a website from which the program can be downloaded are specified.

Cervus

Overview

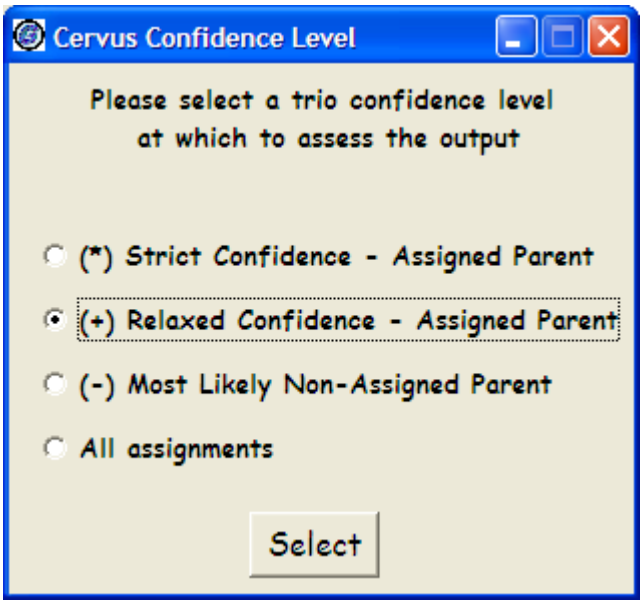
CERVUS uses genetic data from unlinked co-dominant genetic markers such as microsatellites (STRs), single nucleotide polymorphisms (SNP's) and allozymes, it performs allele frequency analysis, parentage analysis simulations, and parentage analysis.

Additional Requirements

Be sure to specify the program output file with the '.csv' extension, and not the text file with the same name.

In order to make the comparison PedAgree requires that all of the candidate parent and offspring files used to create the CERVUS output file be specified in the 'Candidate Parent and Offspring File(s)' area.

When running a CERVUS comparison, you will be prompted by the below form for a level of confidence at which to conduct the analysis.


Cervus Confidence Level
Please select a trio confidence level
at which to assess the output
(*) Strict Confidence - Assigned Parent
(+) Relaxed Confidence - Assigned Parent
(-) Most Likely Non-Assigned Parent
All assignments
Select

The strict and relaxed confidence levels are used as a means of designating assigned parents where the probability is above those levels when derived from the simulations. The default values for strict and relaxed are 95% and 80%

respectively, but any values are able to be specified during the analysis in CERVUS. The most likely non-assigned parent designates parents that would be assigned but their probability falls below the relaxed confidence specification. All assignments includes the rest of the allocated parents not falling into and of the previous categories.

If specifying a CERVUS comparison in a batch file, these confidence levels are specified by 0, 1, 2, and 3 where strict equals 0, relaxed equals 1, most likely equals 2, and all assignments equals 3.

Reference

(Marshall *et al.* 1998,Kalinowski *et al.* 2007)

Download

http://www.fieldgenetics.com/pages/home.jsp

Colony v2.0

Overview

Colony is a computer program implementing a maximum likelihood method to assign/infer parentage and sibship among individuals using their multi-locus genotypes. It can be used in estimating full- and half-sib relationships, assigning parentage, inferring mating systems (polygamous / monogamous) and reproductive skew in both diploid and haplo-diploid species.

In brief, the method (model) assumes a sample of individuals subdivided into 3 sub-samples: offspring (OFS), candidate males (CMS) and candidate females (CFS). Individuals in OFS are assigned (clustered) to K1 paternal and K2 maternal families (K1 and K2 unknown), individuals in CMS and CFS are assigned paternity and maternity to these K1 and K2 families. It is assumed that offspring are either full sibs (sharing both parents), half sibs (sharing only one of the two parents), or unrelated (sharing no parents), while candidates are unrelated and are either parents of or unrelated to the offspring. Markers are

assumed to be in linkage equilibrium and Hardy-Weinberg equilibrium. Violation of these assumptions may lower the power of the analysis, but could be compensated by using more informative markers (Wang 2004). For example, the information about the sex and age of the sampled individuals might be unavailable. In such a case, each individual is allowed to appear in all 3 sub-samples and the sibship and parentage are still inferred satisfactorily in some cases.

Additional Requirements

For parentage analysis select the file with the extension '.BestConfig' extension

Reference

(Wang & Santure 2009)

Download

http://www.zsl.org/science/research/software/colony,1154,AR.html

Famoz

Overview

FAMOZ is used in reconstructing parentage for dominant, codominant and uniparentally inherited markers. Parameters and assumptions used in the calculations are few and simple. Exclusion and identity probabilities, log-likelihoods of any genetic relationship, potential father and parent or parent pair, half- and full-sibship are calculated based on real or simulated data. Error rates for genotypic mistyping can be introduced. Simulations can be done to build statistical tests for parentage assignment.

Additional Requirements

Before running parentage analysis in FAMOZ, clear the output window. Run the analysis using the 'Best Parents/Couples' option under the 'LOD Scores' menu. Once the analysis has been run save the output to a text file.

In order to make the comparison PedAgree requires an ID linking file.

Reference

(Gerber *et al.* 2003)

Download

http://www.pierroton.inra.fr/genetics/labo/Software/Famoz/famozinstall/famozinstall_ct.html

Famsphere

Overview

FAMSPHERE implements a distance-based method to allocate families based on microsatellite marker information. It formalizes the exclusion method considering genotyping errors or mutations. Contrary to the standard exclusion method, FAMSPHERE provides a way to resolve conflicts in the case of multiple paternities. The information linked to the defined distances is similar to that of the transmission probabilities used in the likelihood approaches. However, in FAMSPHERE the information of loci is used as a whole, contrary to the likelihood approaches.

Additional Requirements

Select the file ‘fam.out’ for the program output file.

In order to make the comparison PedAgree requires that both of the candidate parent and offspring files used to create the FAMSPHERE output file be specified in the ‘Candidate Parent and Offspring File(s)’ area.

Reference

(Carvajal-Rodriguez 2007)

Download

http://webs.uvigo.es/acraaj/famsphere.htm

Gimlet

Overview

GIMLET estimates error rates during genotyping and constructing consensus genotypes from repeated genotyping, pools identical genotypes among a set of several genotypes, identifies one (or several) genotype(s) comparing it (them) with references, determines kinship between individuals, and estimates several parameters (allelic frequencies, heterozygosity, probability of identity, population size) from genotyped samples.

Additional Requirements

Make sure to run the ‘Kinship’ option under the ‘Identification’ menu in GIMLET. Specify the ‘pair’ option before running. Once the run has completed, select the file with the ‘(2P).txt’ extension for the program output file.

Reference

(Valiere 2002)

Download

http://pbil.univ-lyon1.fr/software/Gimlet/gimlet%20frame1.html

Newpat

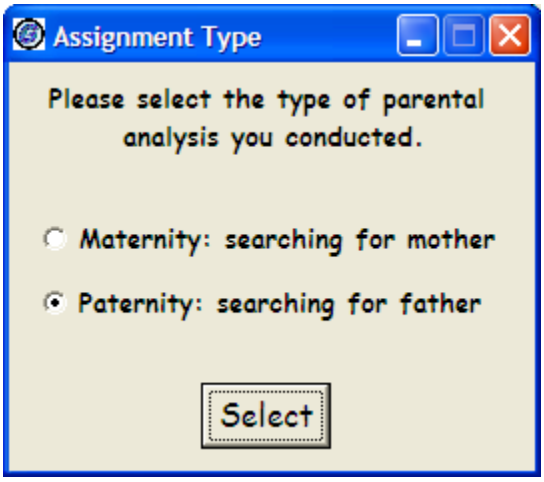
Overview

NEWPAT is a generalized paternity program which calculates allele frequencies, checks for the presence of non-amplifying alleles, assays each input file for duplicate entries, searches for parent-offspring relationships according to user-inputted criteria and then uses a randomization approach to assess the significance of any matches found.

Additional Requirements

Select the 'RESULTS.TXT' for the program output file.

When running a NEWPAT comparison in PedAgree, you will be prompted by the below form to determine if you are conducting a maternity or paternity analysis.


Assignment Type
Please select the type of parental analysis you conducted.
Maternity: searching for mother
Paternity: searching for father
Select

If specifying a NEWPAT comparison in a batch file, a value should be entered on the fourth line. This value should equal 0 if a maternity analysis is being conducted, or 1 if a paternity analysis is being conducted.

Reference

(Wilmer et al. 1999)

Download

http://www.zoo.cam.ac.uk/zoostaff/amos/newpat.htm

Papa

Overview

PAPA is a parental pair allocation and simulator program. The allocation method is based on the likelihood of a parental pair producing the multilocus genotype found in the offspring being tested, which will be referred to as the breeding

likelihood. Estimated level and structure of allele transmission errors in offspring are parameters fed into the allocation procedure. The embodied Monte-Carlo simulator also allows modeling of many allocation conditions, including transmission error and the estimated proportion of missing parents. Simulations may be run prior to the collection of real parents in order to define the minimal set of loci that is necessary to reach a desired level of allocation success. Post-collection simulations aim at statistically assessing the reliability of non-simulated allocations. Simulations output values for several random variables.

Additional Requirements

Be sure to select the 'allocation results.xls' output file, and not the text file of the same name as the program output file.

Reference

(Duchesne *et al.* 2002)

Download

http://www2.bio.ulaval.ca/louisbernatchez/downloads.htm

Parente

Overview

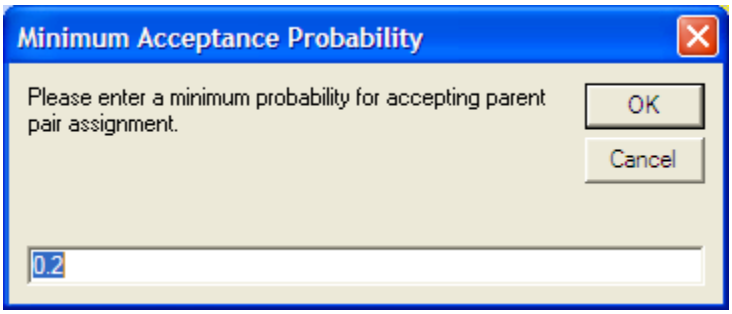
PARENTE uses genotypes of each individual in the sampled population (from codominant markers) and, if available, other individual characteristics (birth and death dates, sex…), to determine the set of potential mothers, potential fathers and potential pairs {mothers; fathers). To determine the set of potential mothers for a given individual, the software first checks for each female whether birth and death dates allow this maternity. Then, it checks for the genetic data compatibility between the individual under consideration and the "age compatible" females. Females (or individuals whose sex is unknown) that satisfy both compatibility conditions are added to the set of potential mothers. The same principle holds for potential fathers. For the pairs {mothers; fathers}, the software checks the

genetic and age compatibilities for all triplets {individual; potential mother, potential father}. In each case, the program calculates the parentage probability.

Additional Requirements

In order to make the comparison PedAgree requires that the candidate parent and offspring file used to create the PARENTE output file be specified in the 'Candidate Parent and Offspring File(s)' area.

When running a PARENTE comparison in PedAgree, you will be prompted by the below form for a probability at which to keep the assigned parent.



This value must be greater than or equal to 0 and less then or equal to 1. If specifying a PARENTE comparison in a batch file, the probability value is entered on the seventh line.

Reference

(Cercueil *et al.* 2002)

Download

http://www-leca.ujf-grenoble.fr/logiciels.htm

Pasos

Overview

PASOS is a parental allocation program designed to identify collected parents based on individual multilocus genotypes while detecting missing parents when a proportion of them have not been collected. It makes use of restricted error tolerance in order to distinguish between a partially incorrect genotype from a false parent's genotype. PASOS also introduces the technique of sequence allocation allowing the user to obtain estimates of the proportion of missing parents and of allocation correctness.

Additional Requirements

Be sure to select the 'allocation results.xls' file for the program output file.

Reference

(Duchesne *et al.* 2005)

Download

http://www2.bio.ulaval.ca/louisbernatchez/downloads.htm

Pedapp

Overview

PEDAPP calculates a detailed relationship structure, typically a pedigree graph or partition, and considers it to be the object of inference. This makes available tools used in complex model selection theory which have demonstrated effectiveness. An important advantage of this approach is that it permits a fully Bayesian approach to the problem, providing a principled and accessible way to measure statistical error. The approach is demonstrated by applying the minimum description length principle. This technique is used in model selection to provide a rational way of comparing models of varying complexity. The resulting score may be interpreted and applied as a Bayesian posterior density.

Additional Requirements

Be sure to select the file containing the '.com' extension for the program output file.

In order to make the comparison PedAgree requires an ID linking file.

Reference

(Almudevar 2007)

Download

http://www.urmc.rochester.edu/biostat/people/faculty/almudevar.cfm

Probmax

Overview

PROBMAX calculates the maximum probability of progeny assignments to a mixture of possible contributing parents, when the genotypes of the parents are known and correspond to the genotypes screened in the progeny, and the parental mating combinations are known (or potentially known).

Additional Requirements

Be sure to select the file containing the '.max' extension for the program output file.

Reference

(Danzmann 1997)

Download

http://www.uoguelph.ca/~rdanzman/software/PROBMAX/

WhichParents

Overview

WHICHPARENTS is a program for determining the most likely parents of offspring using multilocus genotype data. If parental mating history is known, this program also makes use of that information.

Additional Requirements

After the analysis has run in WHICHPARENTS, click anywhere in the report window and select all the output (control + A). Copy the output (control + C) and paste (control + V) it into a new notepad file and save it as your output file.

Reference

(Hedgecock & Eichert 1999)

Download

http://www.bml.ucdavis.edu/whichparents.html

PedAgree

Overview

PedAgree is a program for comparing reconstructed pedigree files to true (known) pedigrees or other reconstructed pedigrees. Additionally, if a sibship reconstruction output file is compared to a parentage reconstruction output file, the two proportionately highest assigned parents for each full-sibling group defined by the sibship reconstruction output file will be tabulated. A pedigree file will be generated using these parents as the parents for all members of each full-sib family that exceed a user-defined threshold.

Additional Requirements

None

Reference

(Coombs *et al.* 2010b)

Download

https://bcrc.bio.umass.edu/pedigreesoftware/

Sibship Reconstruction Output Files

A sibship reconstruction comparison for either the Reconstructed-True or the Reconstructed-Reconstructed Comparison results in the creation of two output files. The first one is a detailed file listing the following information for each individual: Individual identifier, true full-sib family number, true full-sib family size, assigned full-sib family number, and assigned full-sib family size.

Below is an example of a sibship reconstruction details file. There are six true full-sib families, and seven assigned full-sib families. The extra assigned family arose from the splitting of true full-sib family 1. Individual 11479 belongs to true full-sib family 1 which is composed of 22 individuals, but was assigned to full-sib family 2 which contains 11 members.

ID	True Full Sib Family	True Full Sib Family Size	Assigned Full Sib Family	Assigned Full Sib Family Size
11479	1	22	2	11
11481	1	22	1	11
11487	1	22	1	11
11488	1	22	2	11
11489	1	22	2	11
11490	1	22	1	11
11491	1	22	1	11
11493	1	22	1	11
11497	1	22	1	11
11504	1	22	1	11
11507	2	3	3	3
11508	2	3	3	3
11514	3	3	4	3
11520	6	1	7	1
11524	4	2	5	2
11482	1	22	1	11
11485	1	22	2	11
11498	1	22	2	11
11501	1	22	2	11
11502	1	22	2	11
11503	1	22	1	11
11516	3	3	4	3
11476	1	22	2	11

11483	1	22	1	11
11511	2	3	3	3
11517	3	3	4	3
11523	5	2	6	2
11526	4	2	5	2
11477	1	22	2	11
11494	1	22	2	11
11495	1	22	1	11
11505	1	22	2	11
11522	5	2	6	2

The second file is a summary file, and contains the following:

1) If the program producing the sibship output was PARENTAGE there will be a line stating the probability level at which the comparison was conducted.

2) The true and assigned number of full-sib families.

3) The number, size, and composition of each true and assigned full-sib family.

4) The number and proportion of assigned full-sib families required to account for the composition of each true full-sib family.

5) The score reported as the ratio and proportion of correctly assigned individuals out of total individuals where correctly assigned individuals is equal to the total number of individuals minus the minimum number of individuals that must be moved to convert assigned family structure to true family structure.

6) Accuracy of families greater than or equal to a specific family size, and reported for values of 2, 3, 4, 5, and 10. This last section is not a measure of true accuracy in that it only assesses whether individuals assigned to a full-sib family are indeed full-sibs. It does not take into account splitting of larger true full-sib families. For example if a true full-sib family of size ten were assigned to two full-sib families each of size five, they would still be correct in that all members within each assigned family are indeed full-sibs. This error would instead show up in the total accuracy mentioned above.

Below is an example of a sibship reconstruction summary output file. There were nine assigned full-sib families compared to ten true full-sib families. The score for this comparison is 16/20 or 0.8000. This means that four individuals were assigned to the wrong family.

Number of Full Sib Families

True	Assigned
10	9

True Full Sib Families

Family	Size	Member ID's				
1)	5	5034	5037	5040	5044	5046
2)	3	5061	5063	5065		
3)	3	5082	5083	5085		
4)	2	5117	5123			
5)	2	5130	5139			
6)	1	5051				
7)	1	5078				
8)	1	5088				
9)	1	5108				
10)	1	5124				

Assigned Full Sib Families

Family	Size	Member ID's			
1)	4	5051	5082	5083	5085
2)	3	5061	5063	5065	
3)	3	5034	5044	5046	
4)	3	5037	5117	5123	
5)	2	5078	5108		
6)	2	5130	5139		
7)	1	5088			
8)	1	5040			
9)	1	5124			

True Family (Size)	Assigned Family (# in True Family/Size)
1 (5)	3 (3/3)
	4 (1/3)
	8 (1/1)
2 (3)	2 (3/3)
3 (3)	1 (3/4)
4 (2)	4 (2/3)
5 (2)	6 (2/2)
6 (1)	1 (1/4)
7 (1)	5 (1/2)
8 (1)	7 (1/1)
9 (1)	5 (1/2)
10 (1)	9 (1/1)

Total 16 / 20 = 0.8000

Accuracy for assigned full sib families of size >= 10
0 / 0 = 0.0000

Accuracy for assigned full sib families of size >= 5
0 / 0 = 0.0000

Accuracy for assigned full sib families of size >= 4
3 / 4 = 0.7500

Accuracy for assigned full sib families of size >= 3
11 / 13 = 0.8462

Accuracy for assigned full sib families of size >= 2
14 / 17 = 0.8235

Parentage Assignment Output Files

A parentage assignment comparison for either the Reconstructed-True or the Reconstructed-Reconstructed Comparison results in the creation of two output files. The first one is a detailed file that lists the following information for each individual: Reconstructed-True Comparison: Individual identifier, dam identifier, assigned parent 1, sire identifier, assigned parent 2, parent 1 assignment description, parent 2 assignment description, score, misassigned parent 1 relatedness, and misassigned parent 2 relatedness.

The dam and sire identifiers come from the true pedigree file.

The assigned parent 1 and assigned parent 2 identifiers come from the program output file.

The parent 1 and 2 assignment descriptions describe the comparison between the assigned parent 1 and dam identifiers, and the assigned parent 2 and sire identifiers. There are four possible descriptions for a comparison:

1) Right-Sampled means that the true parent was sampled or present in the program input file and the assigned parent equals the true parent.

2) Right-Not Sampled means that the true parent was not sampled and no individual was assigned by the program.

3) Wrong-Sampled means that the true parent was sampled but the program either assigned the wrong individual or no individual at all.

4) Wrong-Not Sampled means that the true parent was not sampled but the program assigned an incorrect individual.

Score provides the number of correct comparisons for the individual. A score of two means that both assigned parents equal their associated true parent. A score of 1 means that one of the assigned parents equals one of the true parents, while the other assigned parent is incorrect. A score of 0 means that neither assigned parent equals either true parent.

The misassigned parent 1 and 2 relatedness columns specify the relatedness between the incorrectly assigned parent and the true parent if the true parent was sampled. The relatedness description can be one of four things:

1) Full-Sib means that the assigned parent shared both of its parents with the true parent.

2) Half-Sib means that the assigned parent shared one of its parents with the true parent.

3) Non-Sib means that the assigned parent did not share either of its parents with the true parent.

4) Not Assigned means that the program did not assign a parent even though the true parent was sampled.

If sexes are unavailable for the candidate parents then relatedness is compared for all combinations of true and assigned parents and the highest level of relatedness is retained.

Reconstructed-Reconstructed Comparison: Individual identifier, output 1 assigned parent 1, output 2 assigned parent 1, output 1 assigned parent 2, output 2 assigned parent 2, parent 1 assignment description, parent 2 assignment description, and score.
The output 1 assigned parent 1 and 2 identifiers come from the Reconstructed Output 1 file.
The output 2 assigned parent 1 and 2 identifiers come from the Reconstructed Output 2 file.
The parent 1 and 2 assignment descriptions describe the comparison between the four assigned parent identifiers. There are 2 possible descriptions for a comparison:

1) Right means that the output 1 assigned parent equals the output 2 assigned parent.

4) Wrong means that the output 1 assigned parent does not equal the output 2 assigned parent.

Score provides the number of correct comparisons for the individual. A score of two means that both assigned parents from output 1 equal their associated assigned parents from output 2. A score of 1 means that one of the assigned parents from output

1 equals one of the assigned parents from output 2, while the other assigned parents do not match. A score of 0 means that neither assigned parent from output 1 equals either assigned parent from output 2.

Below is an excerpt from a Reconstructed-True parentage assignment details file. The program correctly assigned both parents for offspring 11695. The program incorrectly assigned both parents for offspring 11719. Since both or the parents were sampled the relatedness of the incorrectly assigned individuals is given. Parent 1 was a half sibling to the dam, while parent 2 was not a sibling of the sire. For offspring 11722 the program correctly identified the sire while incorrectly assigning the full sibling to the dam. Parent 1 was not assigned for offspring 11753 even though the dam was sampled. Therefore, a relatedness of 'Not Assigned' is given. For offspring 10711 parent 1 was incorrectly assigned given that the true parent was not sampled. Parent 2 was correctly unassigned given that the sire was not sampled. For offspring 10726 both parents were correctly unassigned given that both the dam and sire were not sampled.

Offspring	Dam	Assigned Parent 1	Sire	Assigned Parent 2	Parent 1 Assignment Description	Parent 2 Assignment Description	Score	Misassigned Parent 1 Relatedness	Misassigned Parent 2 Relatedness
11695	11172	11172	10327	10327	Right-Sampled	Right-Sampled	2		
11719	11172	10961	10327	11087	Wrong-Sampled	Wrong-Sampled	0	Half Sib	Non Sib
11722	11172	11159	10327	10327	Wrong-Sampled	Right-Sampled	1	Full Sib	
11753	10699	0	10285	10285	Wrong-Sampled	Right-Sampled	1	Not Assigned	
10711	10106	10360	9366	0	Wrong-Not Sampled	Right-Not Sampled	1		
10726	10106	0	9366	0	Right-Not Sampled	Right-Not Sampled	2		

The second parentage assignment output file is a summary file, and contains the following:

Reconstructed-True:

1) If the program producing the output was CERVUS, PARENTE or NEWPAT there will be a line stating the level, probability or type of analysis for which the comparison was conducted.

2) The ratio and proportion of correctly assigned parents for the following three situations:

A) Both parents were absent or not sampled.

B) One parent was sampled while the other parent was not sampled.

C) Both parents were sampled.

Below the three situations is a summary line giving the ratio and proportion of correctly assigned parents

3) The ratio and proportion of the relatedness categories for situations where the true parent was sampled but not assigned. The relatedness categories are further broken down for the scenarios where only one of the true parents was sampled, or where both of the true parents were sampled.

4) The ratio and proportion of correctly assigned parents for all instances for which a parent was assigned even though the true parent may not have been sampled.

5) The ratio and proportion of correctly assigned parents for instances where both parents of an offspring were assigned even though one or both of the true parents may not have been sampled.

6) The ratio and proportion of assignments for instances when the true parent was sampled. This is further broken down into categories of correctly assigned, incorrectly assigned, and incorrectly unassigned.

7) The ratio and proportion of assignments for instances when an assignment was made. This is further broken down into categories of sampled and correctly assigned, sampled and incorrectly assigned, and unsampled and incorrectly assigned.

Below is an example of a parentage assignment summary file.

Both Parents Absent:	34 / 66 = 0.5152
One Parent Sampled/One Parent Absent:	142 / 208 = 0.6827
Both Parents Sampled:	82 / 108 = 0.7593
Summary:	258 / 382 = 0.6754

Relatedness of misassigned individuals for instances where the true parent was sampled

One Parent Sampled/One Parent Absent:

Full Sib: 5 / 41 = 0.1220

Half Sib: 4 / 41 = 0.0976
Non Sib: 18 / 41 = 0.4390
Not Assigned: 14 / 41 = 0.3415

Both Parents Sampled:
Full Sib: 5 / 14 = 0.3571
Half Sib: 1 / 14 = 0.0714
Non Sib: 7 / 14 = 0.5000
Not Assigned: 1 / 14 = 0.0714

Accuracy for instances where at least one parent was assigned, but not necessarily present
157 / 266 = 0.5902

Accuracy for instances where both parents were assigned, but not necessarily present
98 / 170 = 0.5765

Assignment proportions for instances where a parent was sampled
Sampled and Correctly Assigned 157 / 212 = 0.7406
Sampled and Incorrectly Assigned 40 / 212 = 0.1887
Sampled and Incorrectly Unassigned 15 / 212 = 0.0708

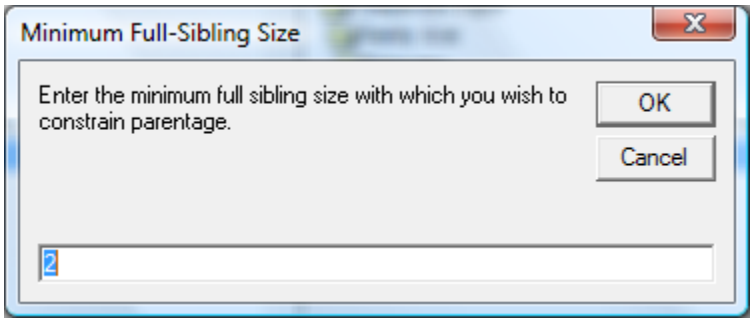
Assignment proportions for instances where an assignment was made
Sampled and Correctly Assigned 157 / 266 = 0.5902
Sampled and Incorrectly Assigned 40 / 266 = 0.1504
Unsampled and Incorrectly Assigned 69 / 266 = 0.2594

Reconstructed-Reconstructed:

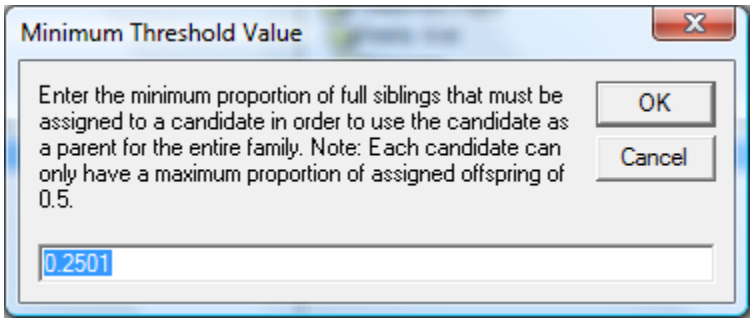
1) If the program producing the output was CERVUS, PARENTE or NEWPAT there will be a line stating the level, probability or type of analysis under which the comparison was conducted.
2) The ratio and proportion of congruence between assigned parents from the two outputs.
3) The ratio and proportion of congruence between assigned parents for varying categories where one or both parents were assigned in one or both outputs.

Sibship-Parentage Comparison

A sibship-parentage reconstruction comparison for the Reconstructed-Reconstructed Comparison option requires user-input for two variables and results in the creation of six output files. The first variable requiring user-input is the Minimum Full-Sibling Size (shown below). This value determines the minimum family size for which to constrain parentage (see below).



The second variable requiring user-input is the Minimum Threshold Value (shown below). This value determines which parents to use for sibship constraint (see below).


Minimum Threshold Value
Enter the minimum proportion of full siblings that must be assigned to a candidate in order to use the candidate as a parent for the entire family. Note: Each candidate can only have a maximum proportion of assigned offspring of 0.5.
OK
Cancel
0.2501

The first output file is a detailed file listing the following information for each individual: Individual identifier, sibship output file full-sib family number, sibship output file full-sib family size, parentage output file full-sib family number, parentage output file full-sib family size, and the parentage output file assigned parents. Below is an example of a parentage-sibship Reconstructed-Reconstructed Comparison details file. Individual '11852' was assigned to full-sib family 1 in the COLONY output file. This family was composed of 53 members. Based on CERVUS parentage assignment, the same individual was assigned to family 2 when parentage was converted to sibship based on shared parents. This family was composed of 24 individuals. Parents '10080' and '11136' were assigned to this individual in the CERVUS output file.

ID	COLONY FS Family	COLONY FS Family Size	CERVUS FS Family	CERVUS FS Family Size	CERVUS Parent 1	CERVUS Parent 2
11852	1	53	2	24	10080	11136

11853	1	53	26	1	10080	11137
11857	1	53	2	24	10080	11136

The second file is a summary file, and contains the following:

1) If the program producing the sibship output was PARENTAGE there will be a line stating the probability level at which the comparison was conducted. If the program producing the parentage output was CERVUS, PARENTE, or NEWPAT there will be a line stating the level, probability, or type of analysis under which the comparison was conducted.

2) The number of full-sib families for the sibship output and the parentage output. The sibship family structure is obtained directly from the sibship output file. The parentage family structure is created from the parentage output file, and is based on the sharing of parents assigned to individuals.

3) The number, size, and composition of each sibship and parentage full-sib family.

4) The number and proportion of parentage full-sib families required to account for the composition of each sibship full-sib family.

5) The score reported as the ratio and proportion of individuals assigned to parentage full-sib families agreeing with sibship full-sib family assignments.

6) The parent pair accounting for the highest proportions of assigned parents for individuals composing sibship full-sib families.

Parts 1 through 5 are identical in appearance to the sibship summary file described above in the sibship reconstruction output files. An example of part 6 is shown below. This file shows that sibship families were acquired from COLONY and parentage was acquired from CERVUS. For family 1, parents ‘11136’ and ‘10080’ accounted for 0.3774 and 0.2736 of individual parent assignments respectively. If all individuals in a full-sib family were assigned the same parent than the proportion for each parent would be 0.5.

COLONY Family	CERVUS Parent 1	Proportion	CERVUS Parent 2	Proportion
1	11136	0.3774	10080	0.2736
2	10931	0.4135	11272	0.2981
3	10989	0.3718	10629	0.2051

Files three, four, and five contain pedigree output that uses the sibship constraint method to improve accuracy of assigned parents. This method uses full-sibling family composition from the sibship reconstruction output file to assess parentage assignment agreement. The method records the proportion of assignments for each candidate within a full-sib family out of the total number of possible assignments. The two candidates with the highest assignment proportions are then compared to the Minimum Threshold Value and, if greater than or equal to that value are assigned to all members of the full-sib family, otherwise an assignment of '0' is given. An example is shown below that corresponds to the data shown for the summary output file above and used a minimum full-sib family size equal to 2, and a minimum threshold value equal to 0.2501. Individuals '11837', '11838', and '11921' were from full-sib family 1 and thus were assigned parents '11136' and '10080' because both of their proportions were greater than 0.2501. Likewise, individuals '12084' and '12040' were from family 2 and thus were assigned parents '10931' and '11272'. However, '12125' and '12149' were only assigned parent '10989' because parent '10629' had a proportion of 0.2051 which is below the minimum threshold value.

Individual ID	Parent 1	Parent 2
11837	11136	10080
11838	11136	10080
11921	11136	10080
12084	10931	11272
12040	10931	11272
12125	10989	0
12149	10989	0

A table showing the number of parentage assignments required to be accepted as parent for entire family based on family size and minimum threshold value is shown below.

Threshold values

Family Size	0.1	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
2					1					2
3			1			2				3
4		1			2		3			4
5		1		2		3		4		5
6	1		2		3	4		5		6
7	1	2		3	4		5	6		7
8	1	2	3		4	5	6	7		8
9	1	2	3	4	5	6	7	8		9
10	1	2	3	4	5	6	7	8	9	10

Output file three is a pedigree file containing individuals and their assigned parents. The pedigree is created by the sibship constraint method using a minimum full-sib family size of one and a minimum threshold value equal to zero. The format of the file is three columns with each row containing the offspring identifier followed by the identifiers of its assigned parents.

Output files four, five, and six produce differing versions of the pedigree file dependent upon the minimum full-sibling size specified. All three output files are formatted identical to file three as described above, with an individual's identifier followed by the identifiers of its two assigned parents. Output file four contains pedigree information for all individuals, but only uses parents constrained by sibship for individuals belonging to full-sibling groups greater than or equal to the user specified value. For individuals belonging to groups smaller than the threshold size, the assigned parents are taken from the parentage assignment output file. The output file has a 'Pedigree-Min FS Size of 5 Complete.txt' extension where '5' changes to equal the user specified value.

Output file five reports an incomplete pedigree containing only individuals belonging to a full-sibling group greater than or equal to the specified threshold. Thus, all individuals in this file have parents assigned using sibship constraint. This file has a 'Pedigree-Min FS Size of 5 Incomplete.txt' extension where '5' changes to equal the user specified value.

Output file six also reports an incomplete pedigree containing only individuals belonging to a full-sibling group greater than or equal to the specified threshold. The difference with this file is that it reports the parents from the source parentage assignment output file, thus allowing for direct comparison to the incomplete, constrained file for accuracy assessment and congruence. Output file six has a ‘Pedigree-Min FS Size of 5 Incomplete.txt’ extension preceded by the name of the parentage assignment program, and where ‘5’ changes to equal the user specified value.

Example Data

Included with the software are a true genotype file, program input and output files for all of the sibship reconstruction and parentage assignment programs, ID linking files and candidate parent and offspring files, both detail and summary comparison PedAgree output files, and a batch file set to run all of the example files. Additionally, there is a file titled ‘Apparent Genotypes.txt’. This file can be used with the program CREATE (Coombs *et al.* 2008) to make new input files for the sibship reconstruction and parentage assignment programs. These files are provided to give the user examples of file formats, output data, and for experimentation with PedAgree.

Disclaimer and Feedback

This program is provided "as-is". Authors and providers are not responsible in case of errors. You may distribute this program freely in any format, so long as the following conditions are met: The program remains intact without modification, the user guide is included without modification, and no fee of any kind is charged.

If there are additional programs that you would like to have added to PedAgree, please let us know. If you encounter any problems with PedAgree please let us know as well. We will attempt to promptly track down and fix any bugs that are reported.

References

Almudevar, A. 2007. A graphical approach to relatedness inference. Theoretical Population Biology **71**:213-229.

Almudevar, A., and C. Field. 1999. Estimation of single-generation sibling relationships based on DNA markers. Journal of Agricultural Biological and Environmental Statistics **4**:136-165.

Ashley, M. V., I. C. Caballero, W. Chaovalitwongse, B. Dasgupta, P. Govindan, S. I. Sheikh, and T. Y. Berger-Wolf. 2009. KINALYZER, a computer program for reconstructing sibling groups. Molecular Ecology Resources **9**:1127-1131.

Butler, K., C. Field, C. M. Herbinger, and B. R. Smith. 2004. Accuracy, efficiency and robustness of four algorithms allowing full sibship reconstruction from DNA marker data. Molecular Ecology **13**:1589-1600.

Carvajal-Rodriguez, A. 2007. FAMSPHERE: a computer program for parental allocation from known genotypic pools. Molecular Ecology Notes **7**:213-216.

Cercueil, A., E. Bellemain, and S. Manel. 2002. PARENTE: Computer program for parentage analysis. Journal of Heredity **93**:458-459.

Coombs, J. A., B. H. Letcher, and K. H. Nislow. 2008. CREATE: a software to create input files from diploid genotypic data for 52 genetic software programs. Molecular Ecology Resources **8**:578-580.

Coombs,J.A., B.H.Letcher, and K.H.Nislow 2010a. PEDAGOG: Software for simulating eco-evolutionary population dynamics. Molecular Ecology Resources in press.

Coombs,J.A., B.H.Letcher, and K.H.Nislow 2010b. PedAgree: Software to quantify error and assess accuracy and congruence for genetically reconstructed pedigree relationships. Conservation Genetics Resources in press.

Danzmann, R. G. 1997. PROBMAX: A computer program for assigning unknown parentage in pedigree analysis from known genotypic pools of parents and progeny. Journal of Heredity **88**:333.

Duchesne, P., T. Castric, and L. Bernatchez. 2005. PASOS (parental allocation of singles in open systems): a computer program for individual parental allocation with missing parents. Molecular Ecology Notes **5**:701-704.

Duchesne, P., M. H. Godbout, and L. Bernatchez. 2002. PAPA (package for the analysis of parental allocation): a computer program for simulated and real parental allocation. Molecular Ecology Notes **2**:191-193.

Emery, A. M., I. J. Wilson, S. Craig, P. R. Boyle, and L. R. Noble. 2001. Assignment of paternity groups without access to parental genotypes: multiple mating and developmental plasticity in squid. Molecular Ecology **10**:1265-1278.

Gerber, S., P. Chabrier, and A. Kremer. 2003. FAMOZ: a software for parentage analysis using dominant, codominant and uniparentally inherited markers. Molecular Ecology Notes **3**:479-481.

Hedgecock D. & Eichert W. WHICHPARENTS (version 1.0): A windows application for determining the most likely parents of offspring using multilocus genotype data. 1999. Computer software distributed from: http://www.bml.ucdavis.edu/whichparents.html.

Kalinowski, S. T., M. L. Taper, and T. C. Marshall. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Molecular Ecology **16**:1099-1106.

Konovalov, D. A., C. Manning, and M. T. Henshaw. 2004. KINGROUP: a program for pedigree relationship reconstruction and kin group assignments using genetic markers. Molecular Ecology Notes **4**:779-782.

Marshall, T. C., J. Slate, L. E. B. Kruuk, and J. M. Pemberton. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology **7**:639-655.

Smith, B. R., C. M. Herbinger, and H. R. Merry. 2001. Accurate partition of individuals into full-sib families from genetic data without parental information. Genetics **158**:1329-1338.

Valiere, N. 2002. GIMLET: a computer program for analysing genetic individual identification data. Molecular Ecology Notes **2**:377-379.

Wang, J., and A. W. Santure. 2009. Parentage and Sibship Inference From Multilocus Genotype Data Under Polygamy. Genetics **181**:1579-1594.

Wang, J. L. 2004. Sibship reconstruction from genetic data with typing errors. Genetics **166**:1963-1979.

Wilmer, J. W., P. J. Allen, P. P. Pomeroy, S. D. Twiss, and W. Amos. 1999. Where have all the fathers gone? An extensive microsatellite analysis of paternity in the grey seal (Halichoerus grypus). Molecular Ecology **8**:1417-1429.